

Using Wikipedia to forecast diseases

November 13, 2014



Los Alamos research published in Public Library of Science

LOS ALAMOS, N.M., Nov. 13, 2014—Scientists can now monitor and forecast diseases around the globe more effectively by analyzing views of Wikipedia articles, according to a team from Los Alamos National Laboratory.

“A global disease-forecasting system will improve the way we respond to epidemics,” scientist Sara Del Valle said. “In the same way we check the weather each morning, individuals and public health officials can monitor disease incidence and plan for the future based on today’s forecast.”

Del Valle and her team were able to successfully monitor influenza in the United States, Poland, Japan and Thailand, dengue fever in Brazil and Thailand, and tuberculosis in China and Thailand. The team was also able to forecast all but one of these,

tuberculosis in China, at least 28 days in advance. This means, Del Valle said, that people start searching for disease-related information on Wikipedia before they seek medical attention.

The paper also shows the potential to transfer models across different regions; that is, one can “train” a computer model using public health data in one location and use it in another. For example, researchers could create models using data from Japan to track and forecast disease in Thailand. This is particularly important for countries that do not offer reliable disease data.

“The goal of this research is to build an operational disease monitoring and forecasting system with open data and open source code,” Del Valle said. “This paper shows we can achieve that goal.”

The scientific paper “Global disease monitoring and forecasting with Wikipedia” is in the Public Library of Science journal (*PLoS Computational Biology*). It is by Nick Generous, Geoffrey Fairchild, Alina Deshpande, Sara Del Valle and Reid Priedhorsky, of Los Alamos National Laboratory’s Defense Systems and Analysis Division.

[Global Disease Monitoring and Forecasting with Wikipedia](#)

Scientists Speak

- [Science Radio Cafe](#) Interview with Sara Del Valle Mathematical Epidemiologist, and Nicholas Generous Molecular Biologist, of the Mathematical and Computational Epidemiology Team at Los Alamos National Laboratory
- [Science Friday Interviews](#) Mining Wikipedia Data to Track Disease

Los Alamos National Laboratory

www.lanl.gov

(505) 667-7000

Los Alamos, NM

Operated by Los Alamos National Security, LLC for the Department of Energy's NNSA

